

**Análisis y estimación de variables aleatorias a través del análisis estadístico  
de la temperatura de la tierra y su correlación con las manchas solares  
implementado en la Institución Educativa Lestonnac**

Alberto López Alzate

Universidad Tecnológica de Pereira  
Facultad de Ciencias Básicas  
Maestría en Enseñanza de la Matemática  
Pereira 2018

**Análisis y estimación de variables aleatorias a través del análisis estadístico de  
la temperatura de la tierra y su correlación con las manchas solares  
implementado en la Institución Educativa Lestonnac**

Alberto López Álzate

Director

M.Sc Fernando Mesa

Documento presentado como requisito para optar al título de  
Magister en Enseñanza de la Matemática

Universidad Tecnológica de Pereira  
Facultad de Ciencias Básicas  
Maestría en Enseñanza de la Matemática  
Pereira 2018

Nota de aceptación

---

---

---

---

Jurado

---

Jurado

---

Jurado

## **Dedicatoria**

A Dios y a mi familia.

## **Agradecimientos**

Gracias al profesor Fernando Mesa por brindarme y darme la oportunidad de trabajar con él en este documento. A la maestría por la formación recibida.

## Tabla de Contenido

<b>Capítulo 1</b>	1
1.1 Planteamiento del problema	1
1.2 Justificación	1
1.3 Objetivos de la investigación	2
1.3.1 Objetivo general	2
1.3.2 Objetivos específicos	2
1.4 Estado del Arte	3
1.5 Algo de historia	5
<b>Capítulo 2</b>	7
2.1 Marco Teórico	7
2.2 Definición	7
2.3 Componentes de las series de tiempo	7
2.3.1 Tendencia secular	7
2.3.2 Variación estacional	8
2.3.3 Variación ciclica	8
2.3.4 Variación irregular	8
2.4 Clasificación descriptiva de las series temporales	8
2.5 Procesos estocásticos	9
2.5.1 Procesos estocastico estacionario	9
2.5.2 Ruido blanco	9
2.6 Autocorrelación	10
2.6.1 Función de autocorrelación	10
2.6.2 Función de autocorrelación parcial	10
2.7 Prueba de LSUNG-BOX	11
2.8 Procesos autoregresivos AR(P)	11
2.9 Proceso autoregresivo de orden 1: AR(1)	12
2.10 Proceso autoregresivo de orden 2: AR(2)	14
2.11 Condiciones de estacionaridad	15

2.12	Proceso de medias móviles $MA(q)$ .....	16
2.13	Proceso de medias móviles de orden 1: $MA(q)$ .....	17
2.14	Proceso de media móvil de orden 2: $MA(2)$ .....	18
2.15	Condiciones de invertibilidad .....	19
2.16	Modelo $MA(2)$ .....	19
2.17	Proceso autoregresivo de medias móviles: $ARMA(p,q)$ .....	20
2.18	Proceso autoregresivo de media móvil: $ARMA(1,1)$ .....	21
2.19	Procesos lineales no estacionarios .....	22
2.20	Uso de operadores y polinomios de retraso .....	23
2.21	ANOVA .....	24
2.22	Teorema central del límite .....	28
<b>Capítulo 3</b>	.....	31
3.1	Metodología .....	31
<b>Capítulo 4</b>	.....	33
4.1	Resultados y discusión .....	33
Conclusiones .....		38
Trabajos Futuros .....		39
Bibliografía .....		40
Apendice .....		42

## Índice de tablas

Tabla 1. Datos .....	33
----------------------	----

## Índice de figuras

Figura 1: Tomado de tierrayestrellas.com .....	4
Figura 2: Curva de la variación del número de Manchas Solares.....	5
Figura 3: Variación de Area.....	36
Figura 4: Variación de Temperatura .....	36
Figura 5: Comparación .....	37



## Capítulo 1

### 1.1 Planteamiento del problema

¿Existe una relación entre la cantidad de manchas solares y la temperatura en la tierra?  
¿Podría formularse un modelo que relacione estas 2 variables y pueda pronosticarse la temperatura en la tierra?

La respuesta a las preguntas anteriores nos permitirá conocer un poco mejor la relación que existe entre estas variables y su impacto sobre las posibles variaciones de la temperatura en la tierra, y las correspondientes consecuencias, tales como fenómenos meteorológicos etc.

Utilizando series de tiempo, se desea encontrar la existencia de una correlación entre dichas variables que explique la relación entre la variable *número de manchas solares* con *la temperatura en la tierra*, relación la cual además pueda predecir la temperatura y de esa manera tomar precauciones respecto a las posibles consecuencias de estas variaciones.

### 1.2 Justificación

Aunque la influencia del sol y en particular de las manchas solares sobre el clima en la tierra requiere de una amplia experiencia en campos como la física del plasma, la química atmosférica, la dinámica de fluidos entre otras; es importante hacer el análisis estadístico que relacione las manchas solares con el clima de la tierra y específicamente su temperatura, como un aporte más en la búsqueda de respuestas a la cantidad de interrogantes que se suscitan a la hora de ver la influencia del sol sobre la tierra y las consecuencias que esta puede acarrear, como por ejemplo: factores climáticos que puedan afectar los cultivos ya que estos podrían favorecer la proliferación de plagas, huracanes, inundaciones.

Más allá de los resultados que se puedan tener, fruto del análisis estadístico, cabe resaltar que este documento es importante como material de consulta, en lo que respecta a las series temporales, puesto que se va a establecer la relación entre las variables ya mencionadas y también con respecto al tiempo, para posteriormente plantear un modelo que permita hacer pronósticos.

En este orden de ideas se puede ver el uso de los modelos AR, MA, ARMA Y ARIMA, además de los supuestos que los sustentan.

Lo anterior es importante porque las series temporales pese a su gran importancia, no han sido muy utilizadas y además poco estudiadas, y los documentos que hay disponibles presentan un alto grado de complejidad. Una de las cosas que se pretende es mostrarlo de la forma más sencilla y entendible posible.

### **1.3 Objetivos de la investigación**

#### **1.3.1 Objetivo general**

Realizar la implementación de la enseñanza de estadística a través de un ejemplo concreto que reúna conceptos aunque parezcan en apariencia difíciles de explicar teóricamente, en una institución educativa.

#### **1.3.2 Objetivos específicos:**

- Analizar y emplear los modelos estadísticos para determinar características de algunas variables.
- Identificar las ventajas de la enseñanza de la estadística a través de ejemplos reales y de interés a los estudiantes.
- Comparar los análisis obtenidos en las diferentes implementaciones.

## 1.4 Estado del arte

Es de vital importancia mirar algunos de los estudios existentes sobre las manchas solares y su posible influencia sobre la temperatura de la tierra. Sin embargo se empieza explicando qué son las manchas solares y cómo estos se originan:

Estas son sencillamente zonas del sol que tienen una temperatura menor a la parte que las rodea, por ejemplo, las manchas se encuentran a unos 4000 °F y la zona que los rodea a unos 6000 °F, que es más o menos la temperatura media de la superficie solar, debido a esto las manchas se ven más oscuras que las calientes.

Hay dos causas que producen manchas solares:

- El material en la capa más externa del sol, se ve obligado a moverse por el camino marcado por el campo magnético.
- El campo magnético de la superficie del sol marca complejos caminos conforme el sol avanza en su ciclo de actividad cada 11 años, al principio del ciclo hay pocas manchas y al final se encuentran muchas. El ciclo termina cuando el campo magnético solar vuelve a tomar una forma sencilla y desaparecen las manchas.

Aparentemente las manchas solares no supondrían una baja de temperatura en la tierra porque la diferencia entre la luz que omite el sol cuando hay manchas y cuando no las hay, no es apreciable, además porque la luminosidad a lo largo del ciclo solar de 11 años no asciende a más del 0,1%, sin embargo los investigadores están comenzando a darse cuenta de que estas aparentemente diminutas variaciones ocasionan cambios significativos en el clima de la tierra.



Figura1. Tomado de tierrayestrellas.com

Gerald Mechh del Centro de investigaciones atmosféricas, presentó el siguiente informe: *Cuando los investigadores analizan los datos correspondientes a la temperatura superficial del océano durante los años en que hay más manchas solares, el pacifico tropical con regiones del pacifico muestran una disminución de un grado centígrado, además muestra indicios de incrementos de precipitación en el pacifico así como de presiones a nivel del mar que están por encima de lo normal en latitudes intermedias del pacifico norte y sur las cuales se correlacionan con los picos del ciclo de manchas solares.*

Se ha estudiado la probable conexión entre el mínimo de Maunder, un déficit en la cantidad de manchas solares que duró 70 años ocurrido a finales del siglo XVII y principios del siglo XVIII el cual fue un periodo muy frio, incluso fue llamada la pequeña era de hielo, en la cual Europa y América del Norte estuvieron sometidos a inviernos bastante fríos.

En cuanto al análisis estadístico de la relación entre manchas solares VS temperatura, hay muy poca información. En la existente, se destaca una denominada *“El ciclo de las manchas solares y la precipitación en la Región del Eje Cafetero-Colombia”* en donde se utilizan series temporales y se establece mediante correlaciones, coeficientes de Pearson y Spearman, una relación inversa entre el índice de las manchas solares y la precipitación anual en estaciones ubicadas en la región del Eje Cafetero. Esta relación presenta una alta significación estadística en tres de las cuatro estaciones. Este estudio se hizo con base a registros que fueron tomados por más de 60 años.

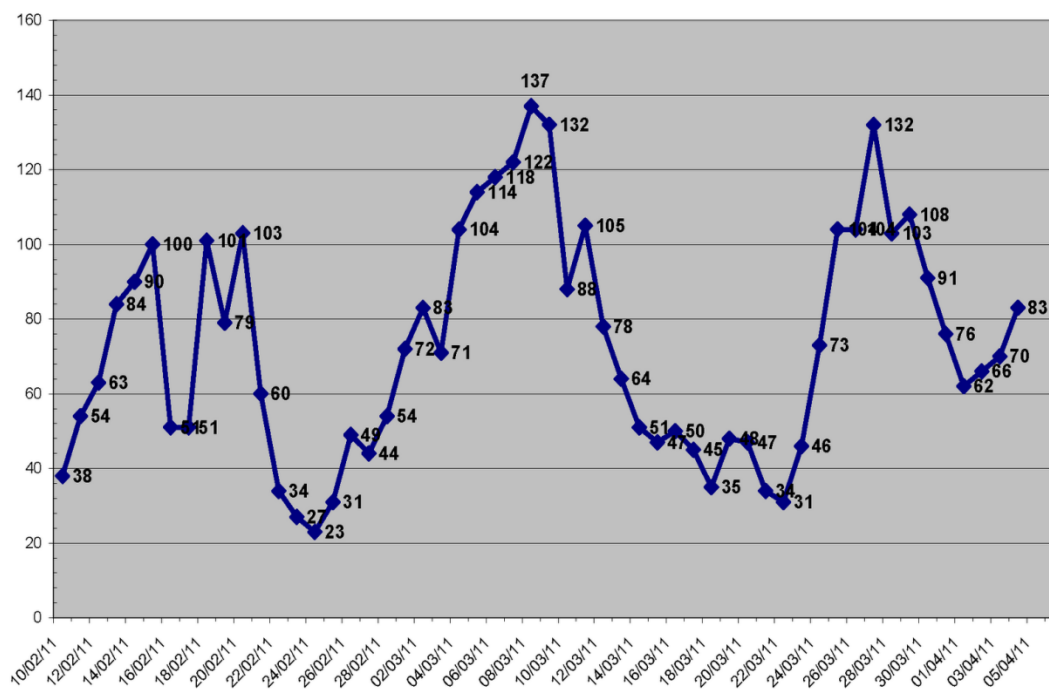


Figura 2. Curva de la variación del Número de Manchas Solares

## 1.5 Algo de historia

El primer registro escrito de las manchas solares fue elaborado por astrónomos chinos alrededor del año 800 A.C. Los astrólogos de la corte de la antigua China y Corea, que creían que las manchas solares eran presagios de eventos importantes, mantuvieron registros esporádicos durante cientos de años de las manchas.

En diciembre de 1128, un monje inglés llamado John Worchester, realizó el primer dibujo de manchas solares poco después de la invención del telescopio, se utilizó para realizar operaciones asociadas con manchas solares. Alrededor del año 1600 los astrónomos no sabían que eran las manchas solares, se pensaba que eran siluetas de planetas cruzando frente al sol.

En el año 1843, el astrónomo alemán Samuel Heinrich Schusabe, descubrió el aumento o disminución de manchas solares.

En abril de 1845 dos físicos franceses, Louis Fixeau y León Foucault, obtuvieron la primera fotografía del sol y las manchas solares

En 1862 cuatro astrónomos notaron que el periodo del ciclo de manchas solares, era idéntico al periodo de cambios en la actividad geomagnética de la tierra, lo que dio origen al campo de estudio de la conexión entre el sol y la tierra que hoy conocemos como “clima espacial”.

Las manchas solares también juegan un papel importante en el clima de la tierra. En el periodo entre 1645 y 1715 se dio el llamado “Mínimo de Maunder” este periodo corresponde al centro, Una serie de inviernos excepcionalmente fríos a lo largo de toda Europa, conocido como “*pequeña edad de hielo*” se continúa debatiendo si esto ocurrió por la poca actividad solar o por otra causa medio ambiental.

Otro periodo entre los años 1.100 y 1250 tenía aparentemente mayor número de manchas solares, las cuales estaban asociadas con mayor actividad solar, este tiempo coincide con un periodo de climas más cálidos en la tierra; a este periodo se le conoce como “periodo medieval cálido”. Desde 1900 el conteo de las manchas solares ha sido mayor, lo cual algunos científicos califican como el *máximo moderno*.

## Capítulo 2

### 2.1 Marco Teórico

**Las series temporales se utilizan cada vez más en diferentes campos, por ejemplo, en economía (tasas de desempleo, tasa de inflación, índice de precios), en meteorología (cantidad de agua caída, temperatura máxima diaria), en geofísica, química, demografía, medicina entre otras.**

### 2.2 Definición

Por series de tiempo nos referimos a datos estadísticos que se recopilan, observan o registran en intervalos normalmente regulares de tiempo (diario, semanal, semestral, anual, etc.)

### 2.3 Componentes de las series de tiempo

Supondremos que en una serie de tiempo existen 4 tipos básicos de variación, los cuales sobrepuestos o actuando en concierto, contribuyen a los cambios observados en un periodo de tiempo, estos 4 componentes son: Tendencia Secular, Variación Estacional, Variación Cíclica, y Variación Irregular.

Supondremos además que existe una relación multiplicativa entre estos cuatro componentes; es decir cualquier valor de una serie es el producto de factores que se pueden atribuir a estos componentes.

#### 2.3.1 Tendencia Secular

Más conocida también como tendencia a largo plazo de una serie, es por lo general, el resultado de factores a largo plazo, la tendencia de una serie de tiempo caracteriza al patrón gradual y consistente de las variaciones de la propia serie, que se consideran consecuencias de fuerzas que afectan el crecimiento o reducción de la misma.

### 2.3.2 Variación Estacional

Representa la variabilidad de los datos debidos a la influencia de las estaciones; corresponde a los movimientos de la serie recurrentes año tras año en los mismos meses (o trimestres) con más o menos la misma intensidad.

### 2.3.3 Variación Cíclica

Con frecuencia las series de tiempo presentan secuencias alternas de puntos abajo y arriba de la línea de tendencia que duran más de un año, esta variación se mantiene después de que se han eliminado las variaciones o tendencias estacionales e irregulares, como los ciclos comerciales, cuyos periodos recurrentes dependen de la prosperidad, recesión, depresión y recuperación, los cuales no dependen de factores como el clima o las costumbres sociales.

### 2.3.4 Variación Irregular

Esta se debe a factores a corto plazo, imprevisibles y no recurrentes que afectan a la serie de tiempo, este componente explica la variabilidad aleatoria de la serie, es impredecible, es decir, no se puede esperar predecir su impacto sobre la serie de tiempo.

## 2.4 Clasificación descriptiva de las series temporales

- a) **Estacionarias:** Si es estable a lo largo del tiempo, es decir, cuando la media y la varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante, y la variabilidad con respecto a esa media también permanece constante en el tiempo.



- b) **No estacionaria:** Son series en las cuales la tendencia y/o variabilidad cambian en el tiempo, los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

## 2.5 Procesos estocásticos

Desde un punto de vista intuitivo, un proceso estocástico se describe como una secuencia de datos que evolucionan en el tiempo. Las series temporales se definen como un caso particular de los procesos estocásticos.

### 2.5.1 Proceso estocástico estacionario

Se dice estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos que dependen únicamente de la distancia o rezago entre estos dos periodos de tiempo y no del tiempo en el cual se ha calculado la covarianza. Sea  $x_t$  una serie de tiempo con estas propiedades:

$$\text{Media: } E(x_t) = E(x_{t+k}) = \mu$$

$$\text{Varianza: } V(x_t) = V(x_{t+k}) = \sigma^2$$

$$\text{Covarianza: } Y_k = E[(x_t - \mu)(x_{t+k} - \mu)]$$

Donde  $Y_k$ , la covarianza (o autocovarianza) al rezago  $k$  es la covarianza entre los valores de  $x_t$  y  $x_{t+k}$  que están separados  $k$  periodos.

Resumiendo, si una serie de tiempo es estacionaria, su media, su varianza y su covarianza (en diferentes rezagos) permanecen iguales sin importar el momento en el cual se midan, es decir son invariantes respecto al tiempo.

### 2.5.2 Ruido blanco

Es un caso simple de los procesos estocásticos donde los valores son independientes e idénticamente distribuidos a lo largo del tiempo con media cero e igual varianza, se denota por  $\varepsilon_t$

$$\varepsilon_t \sim N(0, \sigma^2) \quad \text{cov}(\varepsilon_{t_i}, \varepsilon_{t_j}) = 0 \quad \forall t_i \neq t_j$$

## 2.6 Autocorrelación

En ocasiones en una serie de tiempo acontece que los valores que toma una variable en el tiempo no son independientes entre sí, sino que un valor determinado depende de los valores anteriores, existen dos formas de medir esta dependencia de las variables.

### 2.6.1 Función de autocorrelación (ACF)

La autocorrelación mide la correlación entre 2 variables separadas por  $k$  periodos.

$$\rho_j = \text{corr}(x_j, x_{j-k}) = \frac{\text{cov}(x_j, x_{j-k})}{\sqrt{V(x_j)}\sqrt{V(x_{j-k})}}$$

La función de auto correlación simple tiene las siguientes propiedades:

$$\rho_0 = 1 \quad -1 \leq \rho_j \leq 1 \quad \text{Simetría } \rho_j = \rho_{-j}$$

### 2.6.2 Función de autocorrelación parcial

Mide la correlación entre 2 variables separadas por  $k$  periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas:

$$\pi_j = \text{corr}(x_j, x_{j-k} | x_{j-1} x_{j-2} \dots x_{j-k+1})$$

$$\pi_j = \frac{\text{cov}(x_j - \hat{x}_j; x_{j-k} - \hat{x}_{j-k})}{\sqrt{V(x_j - \hat{x}_j)}\sqrt{V(x_{j-k} - \hat{x}_{j-k})}}$$

## 2.7 Prueba de Isung-Box

Permite probar conjuntamente que todos los coeficientes de autocorrelación son a la vez iguales a 0 , o sea, independientes, se define como:

$$LB = n(n + 2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k} \sim \chi^2(m),$$

Donde  $n$  es el tamaño de la muestra, y  $m$  longitud del rezago

$H_0$ : Las auto correlaciones son independientes

$H_a$ : Las autocorrelaciones no son independientes

En una aplicación si el  $Q$  calculado excede al  $Q$  crítico de la tabla  $Ji$  cuadrada al nivel de significancia seleccionado, no se acepta la hipótesis nula de que todos los coeficientes de autocorrelación son iguales a cero, o por lo menos uno es diferente de 0.

## 2.8 Procesos autoregresivos AR (P)

Se basan en la idea de que el valor actual de la serie  $x_t$ , puede explicarse en función de  $p$  valores pasados  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ , donde  $p$  determina el número de rezagos necesarios para pronosticar el valor actual.

Este modelo de orden  $p$  esta dado por:

$$x_t = \varphi_0 + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \varepsilon_t \quad (1)$$

Que expresado en términos del operador de retardos;

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) x_t = \varepsilon_t$$

Donde  $\varepsilon_t$  es un proceso de ruido blanco y  $\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p$

Son los parámetros del modelo

## 2.9 Proceso Autorregresivo de orden 1 AR(1)

La variable  $x_t$ : está determinada únicamente por el valor pasado, o sea  $x_{t-1}$

$$X_t = \varphi x_{t-1} + \varepsilon_t$$

Con  $\varepsilon_t$  es un proceso de ruido blanco con media 0 y varianza constante  $\sigma^2$ ,  $\varphi$  el parámetro, recordando que en el proceso, el futuro no influye en el pasado.

Prueba de estacionariedad del modelo AR (1) para cualquier valor del parámetro, se debe cumplir:

### a) Estacionario en Media

$$E(x_t) = E(x_t = \varphi x_{t-1} + \varepsilon_t) = \varphi E(x_{t-1})$$

Para que el proceso sea estacionario, la media debe ser constante y finita en el tiempo, o sea:

$$E(x_t) = \varphi E(x_t)$$

$$(1 - \varphi)E(x_t) = 0$$

$$E(x_t) = \frac{0}{1 - \varphi} = 0$$

Entonces para que el proceso sea estacionario el parámetro  $\varphi \neq 1$ .

### b) Estacionario en Covarianza

La varianza tiene que ser constante y finita en el tiempo.

$$\gamma_0 = E(x_t - E(x_t))^2 = E(\varphi x_{t-1} + \varepsilon_t - 0)^2 = \varphi^2 V(x_{t-1}) + \sigma^2$$

Dada la autocorrelación del proceso:

$$E(x_{t-1}\varepsilon_t) = E[(x_{t-1} - 0)(\varepsilon_t - 0)]$$

Si el supuesto es estacionario:

$$E(x_{t-1})^2 = V(x_{t-1}) = V(x_t) = \gamma_0$$

Es decir:

$$\gamma_0 = \varphi\gamma_0 + \sigma^2$$

Entonces:  $\gamma_0 = \frac{\sigma^2}{1-\varphi^2}$

Para que sea estacionario, varianza constante y finita; es necesario que:

$$|\varphi| < 1$$

La función de autocovarianza de orden  $k$  es:

$$\gamma_k = \varphi E(x_t x_{t-k}) + E(\varepsilon_t x_{t-k}) = \varphi\gamma_{k-1}$$

Por lo que:  $\gamma_1 = \varphi\gamma_0, \gamma_2 = \varphi\gamma_1, \gamma_3 = \varphi\gamma_2, \dots$

Por lo tanto se concluye que el proceso  $AR(1)$  es estacionario si y solo si  $|\varphi| < 1$

La función de autocovarianza  $AR(1)$  estacionario es:

$$\gamma_k = \begin{cases} \frac{\sigma^2}{1-\varphi^2} & k = 0 \\ \varphi\gamma_{k-1} & k < 0 \end{cases}$$

Los coeficientes de autocorrelación  $AR(1)$  de un proceso estacionario son:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\varphi\gamma_{k-1}}{\gamma_0} = \varphi\rho_{k-1}$$

La función de autocorrelación de un proceso  $AR(1)$  es:

$$\rho_k = \begin{cases} 1 & k = 0 \\ \varphi\rho_{k-1} & k < 0 \end{cases}$$

Se puede señalar entonces que las características del modelo  $AR(1)$  son:

- El modelo  $AR(1)$  es siempre invertible
- El correlograma representa la gráfica de la función de autocorrelación, este tendrá un comportamiento amortiguado hacia 0 con todos los valores positivos, si  $\varphi > 0$ , y alternando signos si  $\varphi < 0$ .
- La función de autocorrelación parcial se anula para retardos superiores a 1

## 2.10 Proceso autoregresivo de orden 2: $AR(2)$

En los procesos  $AR(2)$  la variable  $x_t$  está determinada por el valor pasado y el anterior a este.

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t \text{ con } \varepsilon_t = \text{ruido blanco}$$

Si hay estacionariedad las características del proceso son:

$$\text{a) } E[(1 - \varphi_1 L - \varphi_2 L^2)x_t] = E[\varepsilon_t]$$

$$(1 - \varphi_1 L - \varphi_2 L^2)E[x_t] = 0 \quad E[x_t] = \frac{0}{1 - \varphi_1 L - \varphi_2 L^2} \quad E[x_T] = 0$$

### b) De Autocovarianza

$\gamma_0 = E(x_t - E[x_t])^2 = E(x_t)^2$ , Se llega a la expresión:

$$\gamma_0 = \frac{\sigma^2 - 2\varphi_1\varphi_2\gamma_1}{1 - \varphi_1^2 - \varphi_2^2}$$

$$\gamma_1 = E(x_t - E[x_t])(x_{t-1} - E[x_{t-1}]) = E(x_t x_{t-1})$$

Que conduce a:  $\gamma_1 = \frac{\varphi_1 \gamma_0}{1 - \varphi_2}$

Donde  $\gamma_1$  y  $\gamma_2$  proporcionan las 2 primeras autocovarianzas en función de los parámetros  $\varphi_1$  y  $\varphi_2$  de la varianza de ruido blanco  $\sigma^2$  las autocovarianzas de orden  $k$ , para todo  $k > 1$  son:

$$\gamma_k = E(x_t - E[x_t])(x_{t-k} - E[x_{t-k}]) = E(x_t x_{t-k})$$

$$\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2}$$

La función de autocovarianza de un modelo  $AR(2)$  es:

$$\gamma_k = \begin{cases} \gamma_0 & k = 0 \\ \gamma_1 & k = 1 \\ \gamma_k & k > 1 \end{cases}$$

**c) Coeficiente De Correlación:**

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} \quad \text{Para } k = 1, 2, 3 \dots$$

Lo que conlleva a escribir como:

$$\rho_k = \begin{cases} 1 & k = 0 \\ \rho_k = \rho_1 \rho_{k-1} + \rho_2 \rho_{k-2} & k > 0 \end{cases}$$

## 2.11 Condiciones de estacionariedad

Modelo  $AR(1)$   $x_t = \varphi x_{t-1} + \varepsilon_t$ , entonces  $(1 - \varphi_L)x_t = \varepsilon_t$

Polinomio autorregresivo  $\varphi_1(L) = 1 - \varphi L$ , pero  $1 - \varphi L = 0$

$$L = \frac{1}{\varphi}$$

Las condiciones de estacionariedad son:

$$L = \left| \frac{1}{\varphi} \right| > 0 \quad |\varphi| < 1$$

Modelo  $AR(2)$

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t \quad (1 - \varphi_1 L^2)x_t = \varepsilon_t$$

Polinomio autorregresivo

$$\varphi_2(L) = 1 - \varphi_1 L - \varphi_2 L^2, \text{ pero } 1 - \varphi_1 L - \varphi_2 L^2 = 0 \text{ sus raíces son } L_1, L_2 = \frac{\varphi_1 \pm \sqrt{\varphi_1^2 + 4\varphi_2}}{-2\varphi_2}$$

Por lo tanto las condiciones de estacionariedad del modelo  $AR(2)$  son:

$$|L_1| = \left| \frac{\varphi_1 + \sqrt{\varphi_1^2 + 4\varphi_2}}{-2\varphi_2} \right| \quad \text{y} \quad |L_2| = \left| \frac{\varphi_1 - \sqrt{\varphi_1^2 + 4\varphi_2}}{-2\varphi_2} \right| > 1$$

Si  $\varphi_1^2 + 4\varphi_2 > 0$  las raíces son reales

Si  $\varphi_1^2 + 4\varphi_2 < 0$  las raíces son complejas

## 2.12 Proceso de medias móviles: $MA(q)$

Modelos determinados por una fuente externa. Estos modelos suponen linealidad, el valor actual de la serie,  $x_t$  esta influenciado por los valores de la fuente externa.

El modelo de promedios móviles de orden  $q$  esta dado por:

$$x_t = \theta_0 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} - \varepsilon_t$$

Expresado en términos del operador de retardos, se tiene que:

$$x_t = (1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q) \varepsilon_t \quad x_t = \theta_q(L) \varepsilon_t$$

Con  $\varepsilon_t$ , un proceso de ruido blanco y  $\theta_1, \theta_2 \dots \theta_q$  los parámetros del modelo



### 2.13 Proceso de medias móviles de orden 1: $MA(1)$

$x_t$  Es función de la innovación actual y su primer retardo, o sea:

$$x_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

Con operador de retardo:

$x_t = (1 - \theta)\varepsilon_t$        $x_t = \theta_1(L)\varepsilon_t$  Con  $\varepsilon_t$  proceso de ruido blanco y  $\theta$  es el Parámetro.

#### a) Estacionario en media

$$E(x_t) = E(\varepsilon_t - \theta \varepsilon_{t-1}) \quad E(x_t) = 0$$

Es estacionario en media, para todo valor del parámetro.

#### b) Estacionario en covarianza

$$\gamma_0 = E(x_t - E(x_t))^2 = E(x_t)^2 = E(\varepsilon_t - \theta \varepsilon_{t-1})^2$$

$\gamma_0 = (1 + \theta^2)\sigma^2 < \infty$  La autocovarianza para  $\gamma_1$  y  $\gamma_2$  es:

$$\gamma_1 = E(x_t - E(x_t))(x_{t-1} - E(x_{t-1})) = E(\varepsilon_t - \theta \varepsilon_{t-1})(\varepsilon_{t-1} - \theta \varepsilon_{t-2})$$

Se llega al siguiente resultado:  $-\theta\sigma^2 < \infty$

$$\gamma_2 = E(x_t - E(x_t))(x_{t-2} - E(x_{t-2})) = E(\varepsilon_t - \theta \varepsilon_{t-1})(\varepsilon_{t-2} - \theta \varepsilon_{t-3})$$

$$\gamma_2 = E(\varepsilon_t \varepsilon_{t-2}) - \theta E(\varepsilon_{t-1} \varepsilon_{t-2}) - \theta E(\varepsilon_t \varepsilon_{t-3}) + \theta^2 E(\varepsilon_{t-1} \varepsilon_{t-3}) = 0,$$

Una forma general de la autocovarianza es:

$$\gamma_2 \begin{cases} \gamma_0 = (1 + \theta^2)\sigma^2 & k=0 \\ \gamma_1 = \theta\sigma^2 & k=1 \\ \gamma_2 = 0 & k > 1 \end{cases}$$

La función de autocovarianza es finita y depende solo de  $K$ , más no del tiempo, para cualquier valor del parámetro  $\theta$  esto implica que no es necesario poner restricciones a  $\theta$

Para que  $MA(1)$  sea estacionario.

La función de autocorrelación de un  $MA(1)$  es :

$$\rho_k = \begin{cases} 1 & k = 0 \\ \frac{-\theta}{1 + \theta^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

## 2.14 Proceso de media móvil de orden 2: $MA(2)$

$$x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

Con  $\theta$ , y  $\theta_2$  y  $\varepsilon_t$  es un proceso de ruido blando, este proceso es estacionario para cualquier valor de  $\theta_1$  y  $\theta_2$ . Sus características son las siguientes:

### a) Estacionario en media

$$E(x_t) = 0$$

### b) Función de autocovarianza :

$$\gamma_k, k = 0, 1, 2, 3 \dots$$

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma^2$$

$$\gamma_1 = (-\theta_1 + \theta_1\theta_2)\sigma^2$$

$$\gamma_2 = (-\theta_2\sigma^2)$$

$$\gamma_3 = 0$$

Resumiendo:

$$\gamma_k = \begin{cases} \gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma^2 & k=0 \\ \gamma_1 = (-\theta_1 + \theta_1\theta_2)\sigma^2 & k=1 \\ \gamma_2 = -\theta_2\sigma^2 & k=2 \\ \gamma_3 = 0 & k>2 \end{cases}$$

Las funciones de autocorrelación están dadas por:

$$\rho_k = \begin{cases} \rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} & k = 1 \\ \rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} & k = 2 \\ \rho_3 = 0 & k = 3 \end{cases}$$

## 2.15 Condiciones de invertibilidad

$$MA(1): \quad x_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad \text{Entonces} \quad x_t = (1 - \theta L)\varepsilon_t$$

El polinomio de medias móviles dado por:  $\theta_1 L = 1 - \theta L$ , para encontrar las raíces del polinomio, se tiene que resolver la ecuación  $1 - \theta L = 0$ , entonces  $L = \frac{1}{\theta}$ . La condición de invertibilidad para  $MA(1)$  es  $|L| = \left|\frac{1}{\theta}\right| > 1$ , esto es  $|\theta| < 1$ .

## 2.16 Modelo $MA(2)$

$x_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$  Entonces  $x_t = (1 - \theta_1 L - \theta_2 L^2)\varepsilon_t$  el polinomio de medias móviles está dado por  $\theta_2 L = 1 - \theta_1 L - \theta_2 L^2$ ,

De donde:

$$L_1, L_2 = \frac{\theta_1 \pm \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2},$$

Representa las condiciones de invertibilidad para el modelo **MA(2)**, están dadas por  $|L_1| =$

$$\left| \frac{\theta_1 + \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2} \right| > 1 \text{ y } |L_2| = \left| \frac{\theta_1 - \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2} \right| > 1$$

## 2.17 Proceso autoregresivo de medias móviles: **ARMA(p, q)**

Si una serie de tiempo  $x_t$  tiene características de *AR* y de *MA* a la vez, es por consiguiente un proceso *ARMA*. Así,  $x_t$  sigue un proceso *ARMA* ( $p, q$ ) con términos autorregresivos y  $q$  términos de media móvil.

$$X_T = C + \varphi_1 X_{T-1} + \dots + \varphi_p X_{T-p} + \theta_1 \varepsilon_{T-1} + \theta_2 \varepsilon_{T-2} + \dots + \theta_q \varepsilon_{T-q} + \varepsilon_T$$

Con  $\varepsilon_t$  es un proceso de ruido blanco, y  $C, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$

Para un proceso *ARMA* ( $p, q$ ) una condición de estacionariedad es la misma que para un proceso *AR* ( $p$ ) una condición de Invertibilidad, es la misma que para un proceso *MA* ( $q$ ).

El modelo *ARMA* ( $p, q$ ) se puede escribir:

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) x_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t$$

$$\varphi_p(L) x_t = \theta_q(L) \varepsilon_t$$

Donde  $\varphi_p(L)$ : es el polinomio autoregresivo y  $\theta_q(L)$  es el polinomio de medias móviles.

Si el proceso es estacionario, su representación *MA*( $\infty$ ) es:

$$X_t = \frac{\theta_q(L)}{\varphi_p(L)} \varepsilon_t, \text{ entonces } X_t = \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \varphi_3 \varepsilon_{t-3} + \dots$$

Si el proceso es invertible, una representación *AR*( $\infty$ ) es:

$$\frac{\varphi_p(L)}{\theta_q(L)} X_t = \varepsilon_t \text{ entonces } X_t = \varepsilon_t + \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \pi_3 Y_{t-3} + \dots$$

Los pasos de la representación  $MA(\infty)$  y  $AR(\infty)$  están restringidos a depender del vector finito de parámetros del modelo  $ARMA(p, q)$ :  $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ .

Un proceso autorregresivo, de medias móviles  $ARMA(p, q)$  es estacionario si y solo si el modulo de las raíces del polinomio autorregresivo  $\varphi_p(L)$  esta fuera del círculo unitario.

Las condiciones de estacionariedad del módulo  $ARMA(p, q)$  vienen impuestas por la parte autoprogresiva, dado que la parte de medias móviles finita siempre es estacionaria. Parte autorregresiva dado que la parte de medias móviles finita, siempre es estacionaria

Un proceso autorregresivo de medias móviles  $ARMA(p, q)$ , es invertible si y solo si el módulo de las raíces del polinomio de medias móviles  $\theta_q(L)$  está fuera del círculo unitario.

Las condiciones de invertibilidad del modelo  $ARMA(p, q)$  vienen impuestas por la parte de medias móviles, dado que la parte autoregresiva es siempre invertible.

Los modelos  $ARMA(p, q)$  siempre van a compartir las características de los modulos  $AR(p)$   $MA(q)$ , dado que contiene ambas estructuras a la vez. El modelo  $ARMA(p, q)$  tiene media cero, varianza constante y finita y una función de autocorrelación infinita. La función de autocorrelación es infinita decreciendo rápidamente hacia 0.

**2.18 Proceso autorregresivo de media móvil  $ARMA(1, 1)$ :**  $x_t$  se determina en función de su pasado hasta el primer retardo, la innovación contemporánea y el pasado de la innovación hasta el retardo 1

$$x_t = \varphi x_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

Donde  $\varepsilon_t$  sigue un proceso de ruido blanco y  $\varphi$  y  $\theta$  son los parámetros del modelo.

Para comprobar la estacionariedad del modelo se calculan las raíces del polinomio autorregresivo

$$1 - \varphi L = 0, \text{ entonces } |L| = \left| \frac{1}{\varphi} \right| \text{ esto es } |\varphi| < 1$$

Para la comprobar la condición de invertibilidad del modelo se calculan las raíces del polinomio de medias móviles

$$1 - \theta L = 0, \text{ entonces } |L| = \left| \frac{1}{\theta} \right| \text{ esto es } |\theta| < 1$$

### Características de $ARMA(1, 1)$ estacionario

#### a) Media

$$E(x_t) = E(\varphi x_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}) = \varphi E(x_{t-1})$$

$$E(x_t) = 0$$

#### b) Función De Autocovarianzas

$$\gamma_0 = E(x_t - E(x_t))^2 = E(x_t)^2$$

$$\gamma_0 = E(\varphi x_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})^2$$

$$\gamma_0 = \frac{(1 - \theta^2 - 2\varphi\theta)\sigma^2}{1 - \varphi^2}$$

$$\gamma_1 = \varphi\gamma_0 - \theta\sigma^2$$

$$\gamma_1 = \varphi\gamma_1$$

Resumiendo los autocovarianzas de un  $ARMA(1, 1)$

$$\gamma_k \begin{cases} \gamma_0 = \frac{(1 - \theta^2 - 2\varphi\theta)\sigma^2}{1 - \varphi^2} & k = 0 \\ \gamma_1 = \varphi\gamma_0 - \theta\sigma^2 & k = 1 \\ \gamma_k = \varphi\gamma_{k-1} & k > 1 \end{cases}$$

La función de autocorrelación de un  $ARMA(1, 1)$  es:

$$\rho_k = \begin{cases} \rho_1 = \varphi - \frac{\theta\sigma^2}{\gamma_0} & k = 0 \\ \rho_k = \varphi\rho_{k-1} & k > 1 \end{cases}$$

## 2.19 Procesos lineales no estacionarios

### Proceso Autorregresivo Integrado Y Media Móvil $ARIMA(p, d, q)$

Los modelos de series de tiempo analizados hasta ahora se basan en el supuesto de estacionariedad. O sea, la media y la varianza para una serie de tiempo son constantes en el

tiempo, y la covarianza es invariante en el tiempo, pero se sabe que muchas series de tiempo no son estacionarias, a este tipo se le considera procesos integrados, se debe diferenciar una serie de tiempo  $d$  veces para hacerla estacionaria y luego aplicarle a esta serie diferenciada un modelo  $ARMA(p, q)$ , se dice que la serie original es  $ARIMA(p, d, q)$ , donde  $p$ , denota el número de términos autoregresivos,  $d$  el numero de veces que la serie debe ser diferenciada para hacerla estacionaria y  $q$  el número de términos de la media móvil invertible, se expresa:

$$x_t^d = c + \varphi_1 x_{t-1}^d + \cdots + \varphi_p x_{t-p}^d + \theta_1 \varepsilon_{t-1}^d + \theta_2 \varepsilon_{t-2}^d + \cdots + \theta_q \varepsilon_{t-q}^d + \varepsilon_t^d$$

Que en la forma de polinomio operador de retardos es:

$$\varphi(L)(1 - L)x_t^d = c + \theta(L)\varepsilon_t$$

Donde  $x_t^d$  es la única serie de la diferencias de orden  $d$ ,  $\varepsilon_t^d$  es un proceso de ruido blanco y  $c, \varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$  son los procesos del modelo.

El modelo  $ARIMA(p, d, q)$  se construye de manera iterativa en 4 etapas:

**Identificación:** el objetivo es determinar lo valores  $p, d$ , y  $q$  apropiados, puede haber más de un modelo candidato que pueda describir la serie.

**Estimación:** se realiza inferencia sobre los parámetros.

**Validación:** se realiza contrastes de diagnóstico para validar si el modelo seleccionado se ajusta a los datos, si no es así, mirar el próximo candidato.

**Predicción:** una vez hecha la selección, hacer los pronósticos de valores futuros

## 2.20 Uso de operadores y polinomios de retraso

Se denota con la letra  $L$  este operador se define mediante la relación  $Lx_t = x_{t-1}$  para todo  $t$ .

Por la aplicación sucesiva del operador  $L$  se obtiene:

$$L^2 x_t = L(Lx_t) = x_{(t-2)}$$

$$L^3 x_t = L(L^2 x_t) = x_{(t-3)}$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$L^k = L(L^{k-1} x_t) = x_{t-k}$$

En general la expresión a la que se llega es:

$$L^k x_t = x_{(t-k)} \text{ para } k = 0, 1, 2, \dots, \text{ y toda } t$$

Al multiplicar a  $L^k$  por  $x_t$  se obtiene la variable retrasada  $k$  periodos y como  $L^0 = I$  se obtiene  $L^0 x_t = x_t$  siendo  $I$  el operador identidad.

El operador modifica toda sucesión de valores  $\{x_1, x_2, \dots, x_N\}$  y la transforma en  $\{x_{L-k}, \dots, x_{t-k}, \dots, x_{N-k}\}$ , por tanto la serie que constaba de  $N$  observaciones se reduce a  $N - k$  observaciones al aplicar  $L^k$ .

Otro operador muy usual es el operador diferencial  $\nabla$  se utilizaba para expresar las relaciones del tipo  $Y_t = x_t - x_{t-1}$  para toda  $t$ .

$Y_t$  Puede escribirse como  $Y_t = \nabla x_t$  la relación de  $\nabla$  un  $L$  es la siguiente  $\nabla = 1 - L$  o sea  $\nabla x_t = (1 - L)x_t$  así como se obtuvo la expresión general para  $L^k$ , así también se podría obtener para  $\nabla^k$

$$\nabla^k x_t = \sum_{j=0}^k \frac{k!}{j! (k-j)!} (-1)^j x_{t-j} \text{ para } k = 0, 1, 2, \dots$$

$$\text{o sea } \nabla_t^{kx} = (1 - L)^k x_t$$

## 2.21 ANOVA

La herramienta estadística que sirve para resolver el problema de comparar más de 2 medias es el “análisis de varianza” que se llama así precisamente porque compara la variabilidad de las medias muestrales (a través de la varianza muestral) con la variabilidad de los elementos dentro de la muestra.

El *ANOVA* permite descomponer la variabilidad total en componentes independientes que pueden asignarse a causas distintas como por ejemplo: variabilidad debida a la máquina, o a la materia prima, o debida a los turnos, etc.

### Fundamentos ANOVA

**Modelo:** los valores de las características de las variables para cada tratamiento tendrán una variabilidad entorno a un valor medio. Si  $Y_{ij}$  con  $j$  valor de la muestra  $i$  el tratamiento:

$$y_{ij} = \mu_i + u_{ij}$$



El problema a resolver es cuál de las 2 situaciones siguientes es la que explica mejor los datos.

- Todos los tratamientos son iguales o sea  $u_1 = u_2 = u_3$
- La media de alguno de los tratamientos es diferente al resto.

### Hipótesis Requeridas

Para que se pueda aplicar el análisis de varianza se deben cumplir 3 hipótesis:

- Los datos han de ser independientes o sea haciendo muestreo aleatorio.
- Las poblaciones han de ser normales.
- Las poblaciones han de tener la misma varianza, es decir, heterocedasticidad.

### Contraste ANOVA

Se puede estimar la varianza de la población  $\sigma^2$  a través de los siguientes estimadores:

- Estimar a  $\sigma^2$  de la población a través de la varianza de cada una de las muestras, lo cual se hace ponderando las varianzas muestrales. Si  $k$  es el número de muestras (tratamientos) mi tamaño de la muestra correspondiente al tratamiento  $i$ -ésimo y  $N$  es el número total de datos disponible en las distintas muestras, este estimador denominado “*varianza residual*” se define como:

$$\hat{\sigma}^2 = S_R^2 = \frac{\sum_i^k (n_i - 1) s_i^2}{\sum_i^k (n_i - 1)} = \frac{\sum_i^k (n_i - 1) S_i^2}{N - k}$$

- Estimar  $\sigma^2$  de la población, suponiendo que los tratamientos no tienen ningún efecto (o sea todos tienen la misma media).

En estas condiciones podemos estimar  $\sigma^2$  a través de la varianza de las medias muestrales:

$$\hat{\sigma}^2 = S_T^2 = \frac{\sum_i^k n_i (\bar{y}_i - \bar{y})^2}{k - 1}$$

De este modo si se verifican las hipótesis, ambas estimaciones no podrán diferir mucho. Si la hipótesis es cierta, el cociente estadístico de ambas varianzas muestrales se distribuyen según una  $F$  es decir

$$\frac{S_T^2}{S_R^2} \propto F_{k-1, n-k}$$

**Resumen para realizar el ANOVA:**

- Fijar el nivel de significación para el contraste, por ejemplo  $\alpha = 95\%$ , establecer el contraste de hipótesis:
  - $H_0$ : Los tratamientos son iguales  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$ .
  - $H_1$ : Alguno de los tratamientos es diferente.
- Calcular  $S_R^2$  y  $S_T^2$
- Calcular la estadística de  $\frac{S_T^2}{S_R^2}$
- Calcular  $F_{k-1, n-k}$  para la significancia prefijada. Si  $\frac{S_T^2}{S_R^2} > F_{k-1, n-k}$  es la diferencia entre los tratamientos entonces es estadísticamente significativa con el nivel de significancia  $\alpha$ .

### Tabla ANOVA

Siendo S la suma de los cuadrados se tiene:

$$S_R^2 = \frac{S_R}{N - k} \qquad S_T^2 = \frac{S_t}{k - 1}$$

Si  $S_D$  es la suma de los cuadrados con respecto a la media global el estadístico  $S_D^2$  es también un estimador de  $\sigma^2$  si se cumplen las hipótesis de igualdad de medias:

$$S_D^2 = \frac{\sum_i^k \sum_j^{n_i} (y_{ij} - \bar{y})^2}{N - 1} = \frac{S_D}{N - 1}$$

Se puede comprobar que:

$$S_D = S_R + S_T$$

En cuanto a los grados de libertad:

$$V_D = V_R + V_T, \text{ es decir } N - 1 = (N - k) + (k - 1)$$

$S_D$  Se calcula con la siguiente ecuación:

$$S_D = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2 - N \bar{y}^2$$

Se calcula  $S_D$  y  $S_T$  calculando  $S_R$  por diferencia.

### Análisis de la diferencia entre las medias

#### a) Procedimiento

El análisis de varianza nos indica si alguno de los grupos es distinto, pero no indica cuáles, para averiguarlo se procede así:

1. Construir una distribución de referencia con la *t de student* las diferencias de las medias de *t de student* para cada uno de los grupos si este solapa o no con los otros.
2. Contrastar las diferencias de las medias de todos los pares posibles de medias utilizando la distribución de *t de student* con el inconveniente de que entre más niveles se analicen, la probabilidad de cometer un error tipo 1 aumenta. Por ejemplo si el nivel de confianza es 0,95 y tenemos 3 grupos, el número de comparaciones es 3, entonces la probabilidad de concluir que un grupo es diferente sin que lo sea, es  $1-(0.95)^3=0.143$ .
3. Método de Bonferroni: Se utiliza cuando el número de grupos es grande porque corrige en parte el efecto anterior.
4. Realizar comparaciones múltiples: Proporcionar intervalos de confianza para las diferencias de las medias de todos los pares de grupos. Los más utilizados son:
  - a. Dunnett: Se utiliza cuando se toma uno de los grupos como referencia.
  - b. MCB de HSU: Compara con el grupo bueno (el más alto o el más bajo).
  - c. Fisher LSD
  - d. Tukey

#### **b) Procedimiento de aplicación**

Como en la mayor parte de los procedimientos estadísticos debe comenzarse por representar gráficamente los datos. Si el número de datos por grupo es menor que 10, se recomienda emplear el diagrama de puntos; si es superior a 10 se recomiendan histogramas o diagramas de caja. En esta representación deben buscarse valores atípicos. Si estos valores atípicos no se deben a un error o una causa subsanable (por ejemplo, error de transcripción de datos) debe pensarse en la necesidad de transformar los datos para que cumplan las hipótesis de normalidad. En la Tabla 5 se dan algunas indicaciones de transformaciones recomendadas. De manera general, se pueden emplear las transformaciones de Box Cox.

Si los datos proceden de un fenómeno de tipo binomial, por ejemplo porcentaje de unidades rechazadas, la transformación adecuada es  $y = \arcsin(p)$ . Si proceden de un fenómeno de tipo Poisson, por ejemplo, el número de defectos, la transformación adecuada es  $y = \sqrt{c}$ .

La construcción de la tabla ANOVA y la realización del contraste puede completarse con la construcción de intervalos de confianza para las medias de cada grupo y los contrastes múltiples.

### c) Validación de las hipótesis

Es muy importante tener en cuenta en todo momento que la validez de las conclusiones está supeditada a que las hipótesis realizadas sean ciertas. Estas comprobaciones pueden hacerse analizando los residuos, es decir las diferencias que existen entre lo explicado por el modelo y los valores obtenidos.

$$y_{ij} = \bar{y}_i + e_{ij}$$

Por lo tanto es preciso realizar las siguientes comprobaciones:

- **Independencia de los datos:** En caso de que los datos se hayan producido según patrones temporales, se deben representar los residuos en la secuencia que se obtuvieron y no deben observarse tendencias, rachas, etc.
- **Normalidad de las perturbaciones:** Los residuos deben distribuirse normalmente. Debe representarse en un papel probabilístico.
- **Heterocedasticidad:** Se representan los residuos por grupos tener una dispersión parecida. Ver por ejemplo la Figura 2. Si el número de datos es al mismo para todos los grupos, el ANOVA es bastante robusto frente a esta hipótesis.

## 2.22 Teorema central del límite

Sean  $x_1, x_2, \dots, x_n$   $n$  variables aleatorias con una distribución de probabilidad no especificada y que tienen una media  $\mu$  y una varianza  $\sigma^2$  finita.

El promedio muestral  $\bar{x} = (x_1 + x_2 + \dots + x_n)/n$  tiene una distribución normal uniforme cuando  $n$  tiende a infinito. En otras palabras la variable aleatoria  $(\bar{x} - \mu)/(\sigma/\sqrt{n})$  tiene como límite una distribución normal estándar.

La esencia del teorema central del límite recae en el hecho de que para  $n$  grande, la distribución de  $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$  es, en forma aproximada, normal con norma 0 y desviación estándar 1. Sin importar cuál sea el modelo de probabilidad a partir del que se obtuvo la muestra.

Debe de tenerse en cuenta que si el modelo de probabilidad de la población es semejante a una distribución normal (si es simétrico y existe una concentración relativamente alta, alrededor del eje de simetría) la aproximación a una distribución normal, será buena, aún para  $n$  pequeño. Por otro lado, si el modelo tiene alta asimetría, la aproximación será buena, para  $n$  grande. En muchos casos puede concluirse de forma segura, que la aproximación será buena si  $n > 30$ . Por lo tanto, la variable aleatoria  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ , se emplea para formular inferencias cerca de  $\mu$  cuando se conoce el valor de la varianza poblacional  $\sigma^2$ . La variable  $z$  es  $N(0, 1)$  cuando el muestreo se lleva a cabo sobre una población que tiene una distribución normal y es, en forma aproximada  $N(0, 1)$  para cualquier otro modelo cuando  $n$  es grande.

### **Demostración Del Teorema Central Del Límite**

Se quiere demostrar que la función generadora de momentos de  $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$  tiende a la de una distribución normal estándar, conforme  $n$  tiende a infinito.

Sean:

$$Z_i = \frac{x_i - \mu}{\sigma} \quad i = 1, 2, \dots, n$$

$Y = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$  Dado que:

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \right) = \frac{1}{n} \cdot \frac{1}{\sigma/\sqrt{n}} \sum_{i=1}^n (x_i - \mu) = \frac{1}{n} \cdot \frac{1}{\sigma/\sqrt{n}} (n\bar{x} - n\mu) = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Entonces,

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$$

Se obtiene que la función generadora de momentos de  $Y$  es igual a la función generadora de momentos de  $1/\sqrt{n} \sum_{i=1}^n Z_i$ , pero  $m_{Y(t)} = \left[ m_{Z_i} \left( t/\sqrt{n} \right) \right]^n = \left\{ E \left[ \exp \left( t Z_i / \sqrt{n} \right) \right] \right\}^n$  dado que las  $Z_i$  son variables aleatorias independientes al expandir  $\left( t Z_i / \sqrt{n} \right)$  en una serie de Taylor:

$\exp\left(tz_i/\sqrt{n}\right) = 1 + \frac{t}{\sqrt{n}}z_i + \frac{t^2}{2n}z_i^2 + \frac{t^3}{3!n^{3/2}}z_i^3 + \dots$  Tomando los valores esperados con  $E(z_i) = 0$  y  $Var(Z_i) = 1$   $i = 1, 2, \dots, n$ .

Se tiene:

$$E\left[\exp\left(tz_i/\sqrt{n}\right)\right] = 1 + \frac{t^2}{2n}z_i^2 + \frac{t^3}{3!n^{3/2}}z_i^3 + \dots$$

De acuerdo a lo anterior:

$$\begin{aligned} m_{y(t)} &= \left[1 + \frac{t^2}{2n}z_i^2 + \frac{t^3}{3!n^{3/2}}E[z_i^3] + \dots\right]^n = \left\{1 + \frac{1}{n}\left[\frac{t^2}{2}z_i^2 + \frac{t^3}{3!\sqrt{n}}E[z_i^3] + \dots\right]\right\}^n \\ &= \left(1 + \frac{u}{n}\right)^n \text{ en donde } u = \frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}}E(z_i^3) + \dots \quad (3) \end{aligned}$$

Ahora  $\lim_{n \rightarrow \infty} m_{y(t)} = \lim_{n \rightarrow \infty} \left(1 + \frac{u}{n}\right)^n$ , por definición  $\lim_{n \rightarrow \infty} \left(1 + \frac{u}{n}\right)^n = e^u$  conforme

$u \rightarrow \infty$  Todos los términos en  $u$ , excepto el primero, tiende  $n$  a 0 debido a que todos tiene potencias positivas de  $n$  en sus denominadores por lo tanto, pueda deducirse que:

$$\lim_{n \rightarrow \infty} m_{y(t)} = \exp\left(t^2/2\right)$$

O la distribución límite de  $Y = (\bar{x} - u)/(\sigma/\sqrt{n})$  es la normal estándar para los valores grandes de  $n$ .

## Capítulo 3

### 3.1 Metodología

Los datos de la temperatura fueron tomados de las bases de datos de la Universidad Nacional Autónoma de México, la dificultad se encontró cuando para una misma fecha de acuerdo al sitio existen diferentes temperaturas, por lo cual se decidió utilizar dos tipos de sitios y realizar un promedio en aquellos sitios donde el primero fue en el trópico y el segundo fuera del trópico.

Los datos de las manchas solares fueron tomados del laboratorio Machine Learning Repository y del Solar Influences Data Analysis Center Observatory of Belgium, la dificultad se encontró en la clasificación de las manchas solares ya que de acuerdo a la región donde ocurre la mancha y al tamaño existe una clasificación para ellas. Por ende, se decidió tomar las más grandes en relación a la cara paralela a la tierra.

La adquisición de datos y su manipulación se realizó con el programa Excel con el fin que los estudiantes de la institución educativa Lestonnac los pudieran manipular adecuadamente.

Los conceptos necesarios para que los estudiantes puedan manipular los datos se encuentran en el capítulo anterior, sin embargo, las masas utilizadas fueron las de estadísticas clásicas, los nuevos conceptos como correlación función teorema del límite central, serie de tiempo se desarrollaron a través del problema. Cabe resaltar este último procedimiento ya que conceptos nuevos y de un mayor grado de complejidad hacía difícil enseñarse sin unos conceptos previos que los estudiantes de secundaria no han adquirido todavía, sin embargo, esto no fue razón para que no se desarrollaran, si bien es cierto no se hizo de manera formal se realizó a través de una forma práctica y se concretó el concepto heurístico.



*Foto 1: sesión de clase en la institución educativa*



## Capítulo 4

### 4.1 Resultados y discusión

A continuación, se mostrará algunos de los datos, de manera completa se encuentran en los anexos.

Tabla 1. *Datos*

DA	TA	1;	1	96	9
2	1	1	2	1	2
3	1	1	2	1	2
3	1	1	2	1	1
3	1	1	2	1	2
3	1	1	2	1	2
2	1	1	2	1	2
2	1	1	2	1	1
2	1	1	2	1	2
3	1	1	2	1	2
3	1	1	2	1	1
3	1	1	2	1	1
2	1	1	2	1	1
3	1	1	2	1	2
3	1	1	2	1	1
2	1	1	2	1	1
3	1	1	2	1	1
3	1	1	2	1	2
2	1	1	2	1	2
2	1	1	2	1	1
2	1	1	2	1	2
2	1	1	2	1	1
3	1	1	2	1	1

3	1	2	2	1	2
2	1	1	2	1	1
2	1	1	1	1	2
3	1	1	2	1	1
3	1	1	2	1	2
3	1	1	2	1	1
3	1	1	2	1	2
3	1	2	2	1	2
3	1	1	2	1	2
3	1	1	2	1	1
3	1	1	2	1	2
3	1	1	2	1	2
Dia					
Area					
Sumatoria					
Prom					
Temp_Med					
1	2	14	1,75	32,8	
	2				
	1				
	2				
	2				
	2				
	1				
	2				
2	2	10	1,25	32,8	
	1				
	1				
	1				
	2				
	1				
	1				

	1			
3	2	12	1,5	32,8
	2			
	1			
	2			
	1			
	1			
	2			
	1			
4	2	13	1,625	32,2
	1			
	2			
	1			
	2			
	2			
	2			
	1			
5	2	13	1,625	32,8
	2			
	2			
	1			
	1			
	1			
	2			
	2			

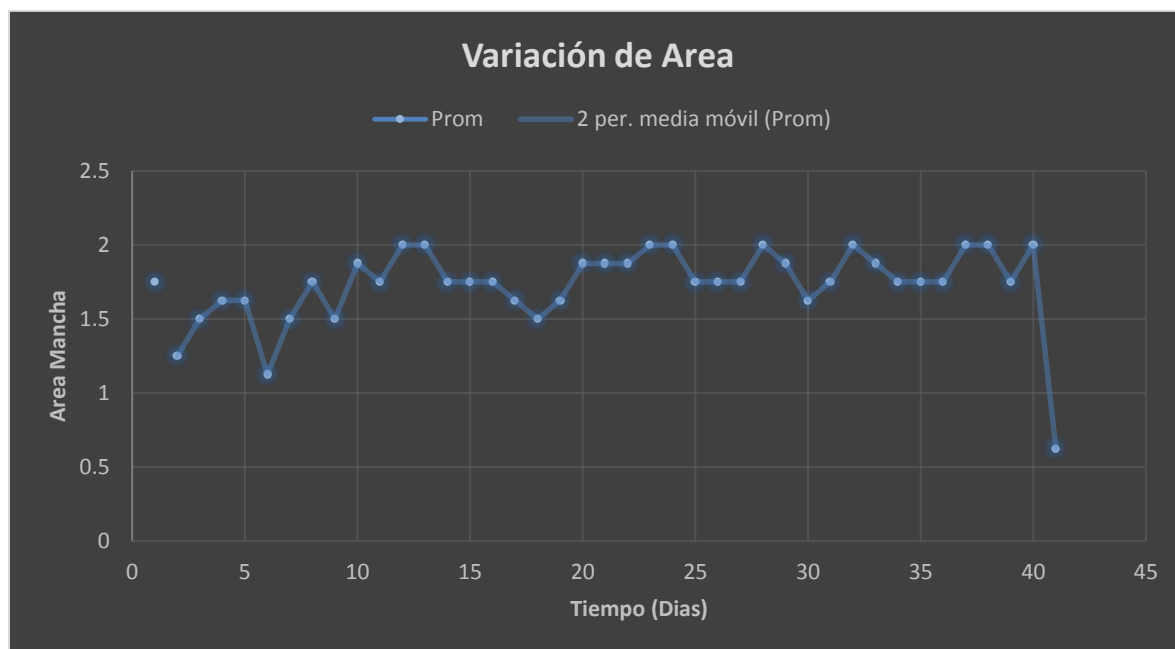


Figura 3: Variación de Area

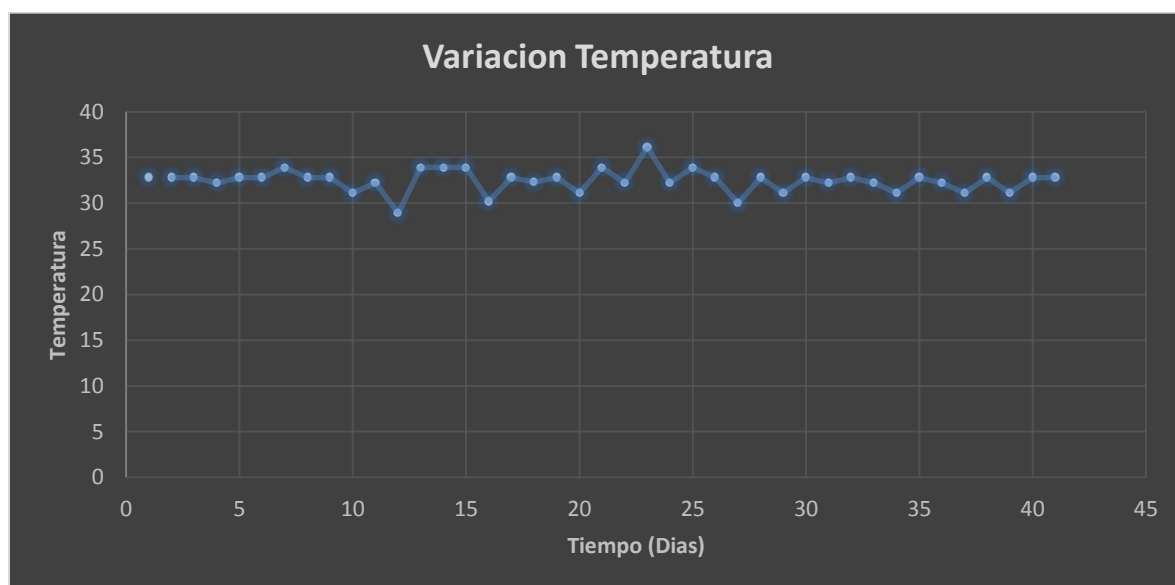


Figura 4: Variación de Temperatura

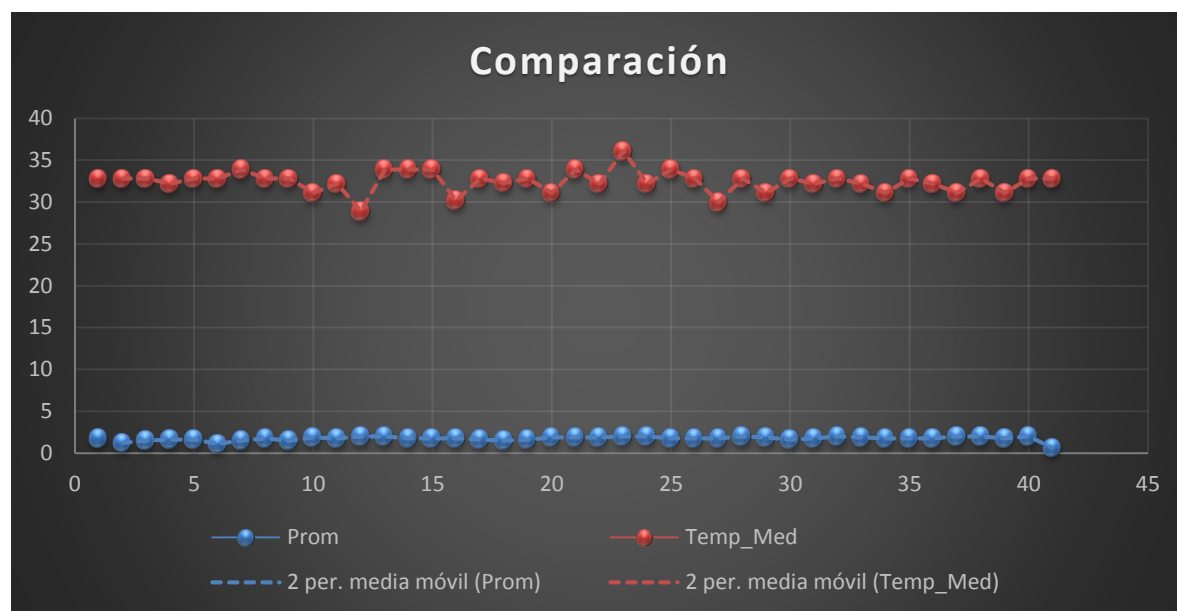


Figura 5: Comparación

## **Conclusiones**

1. Existe una correlación entre las dos variables que nos lleva a pensar en la posibilidad de que el calentamiento global este afectado por la radiación que se genera en la actividad solar. Si bien es cierto que existe dicha correlación esta no es la única variable que debe tenerse en cuenta lo que sería una simplificación ideal.
2. El análisis de datos a través de problemas concretos facilita los procesos de enseñanza de algunos tópicos de la estadística en secundaria.
3. La enseñanza de la estadística como ciencia transversal puede aprovechar las ventajas de abordarla desde ejemplos prácticos y de interés a los estudiantes, ya que en muchas ocasiones, es este el mayor factor que aluden los estudiantes en su proceso de aprendizaje.

### **Trabajos Futuros**

Realizar una correlación con más variables en la que se pueda exponer con mayor realidad el problema.

Utilizar otro tipo de técnicas para realizar el mismo estudio y comparar las diferentes técnicas.

Realizar promedios de temperatura en la tierra en muchos más puntos ya que de acuerdo al sitio la temperatura varía.

Realizar las simulaciones con otro tipo de programas informáticos.

## Bibliografía

- [1] Actividad solar: <http://sidc.oma.be/index.php3> y <http://www.swpc.noaa.gov>
- [2] SALCEDO, Octavio. HERNANDEZ, Cesar A. PEDRAZA, Luis F. Análisis comparativo de las técnicas de series de tiempo ARIMA y ANFIS para pronosticar tráfico Wimax. En: Revista Ingeniería. Vol. 12, No. 2 (Julio, 2007).
- [3] Predicción del ciclo solar: <http://solarscience.msfc.nasa.gov/predict.shtml>
- [4] BENCARDINO, Ciro. Estadística y Muestreo. Colombia: Ecoe Ediciones Ltda., 2000. 918 p. ISBN 9586482448.
- [5] G.G Victor. Manuel, “análisis estadístico de series de tiempo económicas-2a edición”, Editorial: Thomson, 2003.
- [6] W.S. Wei, William, “Time Series Analysis Univariate and Multivariate Methods”, Editorial: Addison Wesley, 1990.
- [7] B. Oksendal, “Stochastic Differential Equations-six edition”, Editorial: Springer, 2006.
- [8] A. Alabert, “Departament de Matemàtiques Universitat Autònoma de Barcelona”, enero 2004. Disponible en:
- <http://mat.uab.es/~alabert/Docs/teaching/IntroSDE%28Cali%29.pdf>.
- [9] C. M. Bencardino, Estadística y muestreo. Ecoe Ediciones, 2000.
- [10] A. Del Sur. (2010, Nov.) Udieconometría e informática. [Online]. Available: [https://www.uam.es/personal\\_pdi/economicas/anadelsur/](https://www.uam.es/personal_pdi/economicas/anadelsur/)



- [11] J. D. Hamilton, Time series analysis. Princeton University Press Princeton, 1994, vol. 2.
- [12] M. y. E. A. IDEAM Instituto de Hidrología. (2015) Boletín climatológico mensual.[Online].Available:<http://www.ideam.gov.co/web/tiempo-y-clima/climatológico-mensual>.
- [13] A. Montanari, R. Rosso, and M. S. Taqqu, “Fractionally differenced arima models applied to hydrologic time series: Identification, estimation, and simulation,” Water resources research, vol. 33, no. 5, pp. 1035–1044, 1997.
- [14] C. Pérez López, “Métodos estadísticos avanzados con spss,” Thompson. Madrid, 2005.
- [15] National Solar Observatory/Sacramento Peak (EEUU), imágenes y datos diarios: [http://nsosp.nso.edu/data/latest\\_solar\\_images.html](http://nsosp.nso.edu/data/latest_solar_images.html)

## Apéndice

### Datos de las manchas solares y de la temperatura: DATA (02/13/69 to 03/27/69)

Dia	Area	Sumatoria	Prom	Temp_Med
1	2	14	1,75	32,8
	2			
	1			
	2			
	2			
	2			
	1			
	2			
2	2	10	1,25	32,8
	1			
	1			
	1			
	2			
	1			
	1			
	1			
3	2	12	1,5	32,8
	2			
	1			
	2			
	1			
	1			
	2			
	1			

4	2	13	1,625	32,2
	1			
	2			
	1			
	2			
	2			
	2			
	1			
5	2	13	1,625	32,8
	2			
	2			
	1			
	1			
	1			
	2			
	2			
6	1	9	1,125	32,8
	1			
	1			
	1			
	1			
	2			
	1			
	1			
7	1	12	1,5	33,9
	2			
	1			
	1			
	2			

	2			
	2			
	1			
8	2	14	1,75	32,8
	1			
	2			
	2			
	1			
	2			
	2			
	2			
9	2	12	1,5	32,8
	2			
	2			
	1			
	2			
	1			
	1			
	1			
10	2	15	1,875	31,1
	2			
	2			
	2			
	1			
	2			
	2			
	2			
11	2	14	1,75	32,2
	2			

	1			
	2			
	2			
	2			
	2			
	1			
12	2	16	2	28,9
	2			
	2			
	2			
	2			
	2			
	2			
	2			
13	2	16	2	33,9
	2			
	2			
	2			
	2			
	2			
	2			
	2			
14	2	14	1,75	33,9
	2			
	2			
	2			
	2			
	2			
	2			
	1			

	1			
15	2	14	1,75	33,9
	2			
	2			
	2			
	2			
	2			
	2			
	1			
	1			
16	2	14	1,75	30,2
	2			
	2			
	1			
	2			
	2			
	1			
	2			
17	2	13	1,625	32,8
	2			
	2			
	1			
	1			
	1			
	2			
	2			
18	2	12	1,5	32,3
	2			
	1			
	1			

	1			
	1			
	2			
	2			
19	2	13	1,625	32,8
	1			
	1			
	1			
	2			
	2			
	2			
	2			
20	2	15	1,875	31,1
	2			
	2			
	2			
	2			
	2			
	1			
	2			
21	2	15	1,875	33,9
	2			
	2			
	1			
	2			
	2			
	2			
	2			
22	2	15	1,875	32,2

	2			
	2			
	1			
	2			
	2			
	2			
	2			
23	2	16	2	36,1
	2			
	2			
	2			
	2			
	2			
	2			
24	2	16	2	32,2
	2			
	2			
	2			
	2			
	2			
	2			
25	2	14	1,75	33,9
	2			
	2			
	1			
	2			
	1			



	2			
	2			
26	2	14	1,75	32,8
	2			
	1			
	1			
	2			
	2			
	2			
	2			
27	1	14	1,75	30
	2			
	2			
	2			
	2			
	2			
	2			
	1			
28	2	16	2	32,8
	2			
	2			
	2			
	2			
	2			
	2			
	2			
29	2	15	1,875	31,1
	2			
	2			



33	2	15	1,875	32,2
	2			
	2			
	2			
	1			
	2			
	2			
	2			
34	2	14	1,75	31,1
	2			
	2			
	2			
	2			
	1			
	1			
	2			
35	2	14	1,75	32,8
	1			
	2			
	2			
	2			
	2			
	2			
	1			
36	1	14	1,75	32,2
	2			
	1			
	2			
	2			

	2			
	2			
	2			
37	2	16	2	31,1
	2			
	2			
	2			
	2			
	2			
	2			
	2			
38	2	16	2	32,8
	2			
	2			
	2			
	2			
	2			
	2			
	2			
39	1	14	1,75	31,1
	1			
	2			
	2			
	2			
	2			
	2			
	2			
40	2	16	2	32,8
	2			

	2			
	2			
	2			
	2			
	2			
	2			
	2			
	2			
41	2			
	1			
	2			
		5	0,625	32,8

